







Rapport d'avancement à mi-parcours

Référence : informatique/projets/promise/nt_ums3365_promise-avancement-mi-parcours

Auteur(s): Franck Gabarrot, Rémy Decoupes

Version du document : va3

Première édition : 01/10/2014 Dernière révision : 23/10/2014

Nombre de pages: 18

Historique des évolutions				
01/10/2014	Création.			
19/10/2014	Mise à jour FG.			
21/10/2014	Mise à jour RD.			
23/10/2014	Intégration du CR du 16/10/2014 et mise en forme finale.			

Abréviations				

Table des matières

1.Documentation.	3
1.1 Documents de référence	3
1.2 Documents associés	3
2.Introduction.	
2.1 Contexte de mise en œuvre	4
2.2 Rappel du projet	
3.Bilan général	
3.1 Organisation du projet	
3.2 Modèle construit pour l'organisation des logiciels	
3.3 Première phase du projet : validation du cahier des charges & échanges (mission #1)	7
3.4 Phase de développement – Promise/lidar : avancement	
3.5 Phase de développement – Promise/web : avancement	
3.6 Phase de développement – Promise/arletty : avancement	
3.7 Statut sur les livrables et la documentation associée	
3.8 Rayonnement du projet	
3.9 Interface station-centre de données et standardisation du format des données station	
4.Conlusion.	

1. DOCUMENTATION

1.1 Documents de référence

Code	Référence	Titre	Auteurs	Date
DR01	informatique/projets/promise/nt_um s3365_promise-demande-cu- ether_20131128	Demande d'utilisation de ressources ETHER Projet PROMISE	Franck Gabarrot	2013/11/28
DR02	informatique/projets/promise/nt_um s3365_promise-adaptation-1- an_20150305	Projet PROMISE : adapation des réalisations sur 1 an	Franck Gabarrot, Philippe Keckhut, Jean-Pierre Cammas	2014/03/05
DR03	CR_promise_20140424	Compte rendu de la réunion de démarrage PROMISE du 24 avril 2014	Cathy Boonne	2014/04/24
DR04	informatique/projets/promise/nt_um s3365_promise-logicielslidar	Cahier des charges techniques Promise/lidar	Franck Gabarrot, Rémy Decoupes	2014/04/30
DR05	informatique/projets/promise/nt_um s3365_promise-logicielsweb	Cahier des charges techniques Promise/web	Rémy Decoupes, Franck Gabarrot	2014/04/30
DR06	informatique/projets/promise/cr_um s3365_promise-mission1_20140516	Compte rendu Mission #1 du 13 au 16 mai 2014	Franck Gabarrot, Rémy Decoupes	2014/05/16
DR07	informatique/projets/promise/nt_um s3365_guide-developpement- logiciels	Guide pour le développement de logiciels	Franck Gabarrot, Rémy Decoupes	2014/07/31 (va0)
DR08	arletty_doc	Document de description de la chaîne de production et de mise à disposition des données Arletty	Cathy Boone	2014/10/14
DR09	informatique/projets/promise/cr_um s3365_avancement- promise_20141016	Compte rendu de la réunion d'avancement PROMISE du 16 octobre 2014	Franck Gabarrot, Cathy Boone	2014/10/16

1.2 Documents associés

Code	Référence	Titre	Auteurs	Date
DAxxx				

2. INTRODUCTION

2.1 Contexte de mise en œuvre

Projet soutenu par le CU Ether (décembre 2013) et validé sous conditions par le CD Ether

Soutien accordé : support budgétaire pour un CDD niveau IR et pour des missions entre le CGTD Ether (Paris) et l'OPAR (Réunion).

Durée du soutien accordé : 1 an (avril 2014 à mars 2015).

Le document DR01 présente le projet initial et décrit son adaptation sur la durée d'un an (réf. DR02).

Responsable réalisation du projet & développeur sur le projet : Franck Gabarrot.

Développeur sur le support CDD accordé : Rémy Decoupes, ISPL-CGTD ETHER, mis à disposition pour la réalisation du projet à l'OSU Réunion. Responsables locaux : Franck Gabarrot (Responsable informatique OSU Réunion/UMS 3365) et Jean-Pierre Cammas (Directeur OSU Réunion/UMS 3365).

2.2 Rappel du projet

Ce projet a pour objectif de répondre aux difficultés d'alimentation de la base de données ETHER-NDACC en données géophysiques validées issues des stations d'observations NDACC-France, et en particulier pour l'OPAR. PROMISE vise à apporter une réponse à ces difficultés en proposant un « refactoring » des outils de traitement, de gestion et de diffusion des données dans les stations afin de les rendre opérationnels et de permettre la reprise d'un flux stable de données. On se base donc sur un existant qui n'est aujourd'hui tout simplement plus maintenable, ce qui est le facteur bloquant majeur pour alimenter la base de données NDACC. PROMISE s'appuie sur l'expertise technique et organisationnelle des acteurs en station. C'est une modernisation des outils stations et des méthodes de travail qui vise à proposer des outils mutualisables et une organisation de projet favorisant les échanges techniques. PROMISE est orienté sur les besoins lidar NDACC-France associés aux stations principales que sont l'OPAR, l'OHP, et DDU. Mais le projet est pensé pour être exploitable pour les autres instruments et pour les autres stations françaises.

Le développement initialement proposé visait le logiciel de gestion des flux de données et le logiciel d'archivage-diffusion-et-validation-en-ligne sur un délai de travail de 2 ans (cf. DR01). Le projet a été revu sur un délai de travail de 1 an, le développement du logiciel de gestion des flux de données a été abandonné et remplacé par le développement d'un framework de traitement lidar jugé plus critique : PROMISE/LIDAR. Le développement du logiciel d'archivage-diffusion-et-validation-en-ligne a été maintenu : PROMISE/WEB. Le projet comprend également une partie assurée par le CGTD ETHER IPSL pour la mise à disposition des données ARLETTY pour chaque station NDACC-France : PROMISE/ARLETTY.

Au delà du développement d'un framework lidar, PROMISE doit permettre de travailler sur des points d'organisation générale qui vont favoriser la stabilité sur le long terme du flux de données : la définition d'un format de données station et les spécifications fonctionnelles et techniques de l'interface d'échanges entre la station et le centre de données.

Le logiciel d'archivage-diffusion-et-validation-en-ligne a pour ambition, non seulement de gérer les données et de faciliter leur envoi vers le centre de données, mais aussi d'apporter des solutions à des difficultés d'organisation du métier : 1) faire de la production et de la gestion de données de façon quasi-opérationnelle dans un environnement recherche en développement constant (=la

station); 2) fournir une plateforme de travail commune à des personnes géographiquement dispersées et ayant des métiers différents (PI et chercheurs associés, acteurs techniques station, équipes extérieures); 3) conserver l'information (descriptif et statut instrumental, conditions d'acquisition, critères de validation, etc).

3. BILAN GÉNÉRAL

3.1 Organisation du projet

- Rémy Découpes : développement logiciel PROMISE/WEB.
- Franck Gabarrot : développement logiciel PROMISE/LIDAR.
- Importance mise sur l'organisation de l'équipe de développement avec des échanges techniques fréquents (point technique une fois/semaine minimum), des méthodes (Agile, DR07) et des outils de gestion des développements (cf. point ci-après).
- Mise en place d'un serveur de développement ouvert basé sur GitLab et MediaWiki : devSpot accessible via http://osur.univ-reunion.fr/outils/devspot/

Pour vous connecter au serveur GitLab, ci-après le login/password invité pour le groupe Promise : promise_guest / promise2014

- Mutualisation de technologies entre les deux logiciels du projet : Python 2.7 comme langage principal, développement d'une librairie métier transverse : MiscBox.
- Choix des technologies en cohérence avec les choix des équipes techniques des autres stations françaises (en particulier Python et le framework web Django).
- Importance mise sur la structuration des paquets logiciels, la documentation et la facilité d'utilisation (ex. : livraison sur un live CD) : pérennité et utilisabilité.
- 2 missions au CGTD Ether : #1 démarrage et validation du cahier des charges et de la conception, #2 recette et déploiement OHP et DDU.

3.2 Modèle construit pour l'organisation des logiciels

Ci-après un extrait du document DR07 :

Documentation associée à un logiciel :

- Document de développement (odt/pdf): conception générale, interfaces, conception détaillée et algorithmie, clés de maintenance ;
- Manuel de réféfence (html): liste de toutes les fonctions, avec leurs structures d'appel, la description des entrées et des sorties (exploitation des commentaires structurés avec Doxygen);
- Document de validation et de suivi (odt/pdf) : les méthodes de validation et les résultats pour chaque version.
- Manuel d'utilisation (odt/pdf): installation et utilisation.

Structuration d'un logiciel en python:

```
PaquetLogiciel/ ←- nom du logiciel
bin/ ←- fichiers exécutables générés
lib/ ←- librairies générées

paquetlogiciel/ ←- fichiers sources (en miniscule) dont __init__.py

config/ ←- fichiers de configuration et/ou de paramétrage, on peut aussi
mettre se répertoire dans le répertoire paquetlogiciel, en python c'est plus logique.
```

```
utils/
           ← - utilitaires générés ou associés
              ← - sources et executables de test
 test/
 doc/
              ← - documentation
                            ← - doc algo de référence
             algorithms/
  development_manual/
                            ← - doc de conception, de développement, schémas, etc
                            ← - doc technique de référence (Doxygen)
       reference_guide/
                     html/
                               ← - dans sa version html
            user guide/
                            ← - doc utilisateur
             validation/
                            ←- doc de validation et de suivi des anomalies
AUTHOR
             ← - fichier contenant le ou les noms des auteurs
                   ← - fichier traçant les modifications
CHANGELOG
                    ← - fichier contenant la licence (+ LICENSE-fr)
LICENSE
README
              ← - description du paquet logiciel (référence à la documentation dans
                  doc s'il y en a). On peut faire un README.md (format ReST)
                  qui est plus pythoniste si on est chaud.
INSTALL
              ←- notes pour l'installation (référence à la documentation utilisateur
    s'il y en a une)
TODO
               ← - qu'est-ce qu'on doit encore faire et/ou prévu de faire
VERSION
               ← - version du paquet x.x.x
                ← - fichier d'installation Python
setup.py
MANIFEST.in ←- fichier texte qui liste les fichiers non python à inclure dans
                     l'installation
```

3.3 Première phase du projet : validation du cahier des charges & échanges (mission #1)

Le premier mois du projet a été consacré à l'élaboration des cahiers des charges techniques des logiciels à développer (cf. DR04 et DR05) ainsi qu'à leurs conceptions. Il s'est conclu par une mission d'une semaine de Franck Gabarrot et de Rémy Decoupes au CGTD Ether IPSL avec pour objectif la validation des cahiers des charges et la présentation détaillée du projet aux différents responsables scientifiques des instruments NDACC du LATMOS (et du LACy en décalé). Le document DR06 fournit un compte rendu détaillé de cette mission.

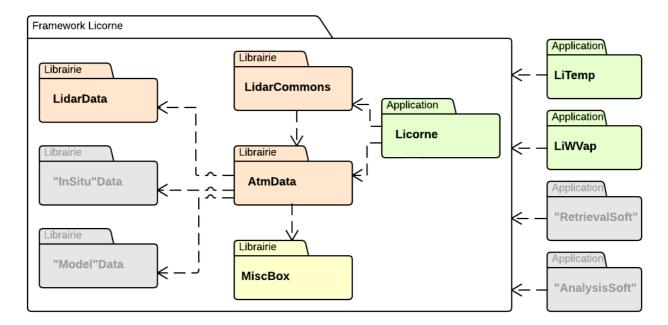
Les échanges ont été très positifs :

- Accueil très favorable du projet et réelle attente de la part des PIs des instruments.
- Echanges avec le CGTD Ether très constructifs avec une volonté forte de faire progresser l'interface d'échanges entre les stations et le CGTD Ether.

3.4 Phase de développement – Promise/lidar : avancement

3.4.1 Description et état d'avancement

Ci-après l'architecture générale du framework lidar **Licorne** (« Lidar processing corner ») développé dans le cadre du projet.



Il s'agit d'un framework mixte C++/Python pour la gestion et le traitement des données lidar. Il a pour objectif dans sa conception de pouvoir être exploiter aussi bien dans un cadre de production automatisé de produits élaborés que dans un cadre de traitement pas à pas. L'application « Licorne » a justement pour objectif de fournir les moyens d'éxecuter de façon unitaire chacune des fonctionnalités d'import/export ou de traitement du framework. La version de l'application « Licorne » élaborée dans le cadre ce projet sera modeste (en ligne de commandes) mais l'ambition future est de développer une interface graphique pour cette application.

MiscBox est une librairie métier entièrement en Python comportant notamment des outils d'exploitation de fichiers XML pour la structuration de métadonnées et pour la génération d'interfaces graphiques de saisie à partir de modèles au format XML. Elle est disponible en version 0.9.

LidarData est une librairie écrite en C++ et possédant une surcouche en Python. Elle permet la lecture écriture de données lidar dans différents formats. Elle fonctionne sous Linux ou Windows. La version 1.0 est disponible. Une version 1.1 incluant d'autres formats et l'utilisation du NetCDF est en cours (la version 1.0 permet l'exploitation du HDF4).

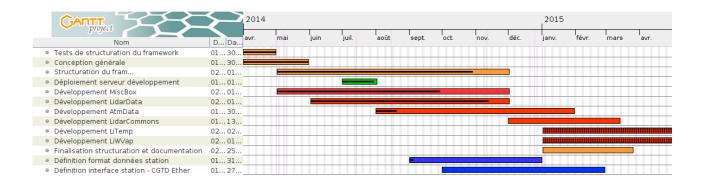
AtmData est également une librairie écrite en C++ avec une surcouche Python. Elle exploite la librairie LidarData. Son objectif est la gestion des données en proposant un conteneur de données générique (GeoData) qui pourra être utilisé pour les données lidar, de GPS (GNSS), de radiosonsage, de modèles, etc. AtmData représente le niveau supérieur d'abstraction du format et de la structuration des données par rapport à LidarData. AtmData doit proposer des fonctions de traitement de base : filtrage, échantillonnage, etc. C'est la librairie centrale du framework et le point le plus difficile du développement de ce dernier. Son développement est en cours, il est en retard par rapport au planning initial (>1 mois).

La libraire LidarCommons sera écrite en Python principalement, elle s'appuiera sur la librairie AtmData qui lui permettra une abstraction du format des données et de la structuration des fichiers. Elle comprendra l'ensemble des fonctions de traitement lidar utiles pour le pré-traitement et l'inversion. Le développement de la branche principale n'a pas démarré. Une version de test – non basée sur AtmData – a été développée (incluant principalement somphot) afin de tester l'organisation du logiciel et surtout de tester la faisabilité d'exploiter le code de traitement de niveau 2 pour la température en fortran 77 (qui a été adapté et transformé pour l'occasion en librairie avec

une surcouche Python). Le développement doit démarrer en décembre 2014 et se finaliser avec une première version contenant les fonctions de base en mars 2015.

Le niveau technique de la réalisation est satisfaisant. L'utilisation mixte de C++ et Python est très performante (utilisation de boost.python, CMake pour l'automatisation de la compilation). La méthode de structuration des paquets logiciels est mature et suit les règles de l'art, ce qui donne un framework à ce stade assez clair.

Le taux de réalisation de ce développement n'est pas complétement satisfaisant et il sera difficile de tenir tous les objectifs à 1 an (voir ci-après le diagramme de Gantt du projet Promise/lidar au 01/10/2014). Néanmoins la dynamique est positive et le choix est fait de favoriser la qualité et la pérennité de ce qui est développé par rapport à la finalisation de tout. Le sous-chapitre qui suit indique les tâches prioritaires sur la fin du projet.



3.4.2 Priorisation des tâches qui restent à mener

Afin d'optimiser le temps de développement restant, le module de tracé graphique de la lirairie AtmData est pour le moment abandonné. On va se focaliser sur l'essentiel : terminer les classes de base de la librairie et intégrer le module de lecture/écriture NetCDF4 − CF accompagné des templates station pour les produits lidar (→ définition d'un standard pour les données station). La généralisation du NetCDF pour les fichiers de sortie permettra d'utiliser des outils existants s'appuyant sur ce format pour la partie graphique.

Pour LidarCommons, le développement des fonctions de traitement de niveau 1 sera prioritaire sur celui des fonctions de niveau 2. L'objectif général reste de livrer une première version exploitable du framework Licorne avec un développement de base solide permettant sa maintenance et son évolution. Il sera donc difficile de mener à bien de façon satisfaisante dans le cadre du projet le développement des applicatifs d'inversion LiTemp et LiWVap.

Etant donné l'avancement du travail sur la filière température et notamment la disponibilité du code de traitement d'Alain Hauchecorne ré-organisé en librairie fortran 77 avec une interface Python, on conserve l'objectif de fournir une première version de LiTemp fin mars 2015 (version 0.9). Il nous manquera juste un peu de recul sur ses performances sur une grande variété de jeu de données.

3.5 Phase de développement – Promise/web : avancement

3.5.1 Description

Le développement de la partie PROMISE/WEB, nommé **e-obs**, est un système d'exploitation modifié afin d'incorporer nativement toutes les fonctionnalités développées dans le cadre du projet. Livré sous forme de live DVD, son installation est, de fait, très simplifiée. E-obs est conçu comme un espace de travail et d'échange pour les différents acteurs de la production des données dans la station et un lien entre la station et le centre de données.

L'élaboration du système e-obs peut être divisé en 3 tâches majeures :

- Développement logiciel :
 - Application Web: répond aux sollicitations des internautes afin d'afficher-modifierajouter des données dans la base de données
 - Application tâche de fond ou démon : gère l'alimentation automatique de la base de données en fonction des données station entrantes ainsi que les envois automatiques vers le CGTD après validation des données par les PI
- Administration Système live DVD: fourni un environnement de production facile d'installation (avec serveur Web, base de données et les différents composants logiciels d'eobs).
- Gestion de projet informatique : offre toute la documentation liée à un projet informatique (cahier des charges, conception, développement, validation et manuel d'utilisateur). A ceci, nous ajoutons aussi les phases de validations fonctionnelles.

En ce qui concerne l'application Web, un gros effort sur l'ergonomie et le design est réalisé. Elle propose plusieurs espaces de travail en fonction du rôle de l'internaute :

- PI : Validation des données et visualisation des modifications sur l'instrument.
- Opérateur : ajout de métadonnées.
- Administrateur système : gestion de la base de données et des utilisateurs.

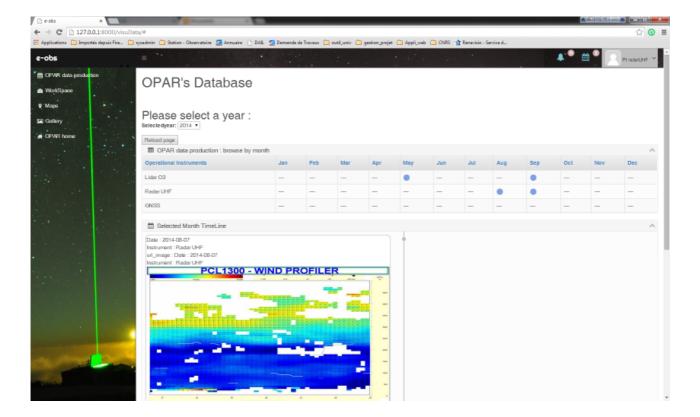
D'autres fonctionnalités communes sont présentes : visualisation de la production en temps quasiréel des données station, géolocalisation des instruments, etc... En effet, grâce à la modularité du système, d'autres fonctionnalités peuvent être facilement ajoutées.

Afin d'obtenir des premiers retours sur notre mode de fonctionnement, nous avons mis à disposition une première version d'e-obs début juillet. Pour cela, nous avons consacré les 3 premiers mois du projet à la documentation, à la mise en place d'une architecture logicielle modulaire avec les premiers développements, ainsi qu'à l'élaboration du liveDvd. Au niveau du développement logiciel, cette livraison incorpore une version de l'application Web, qui est un démonstrateur de ce qu'il sera réalisé par la suite. La livraison a été effectuée le 09/07/2014 et mise à disposition d'ether et des stations OHP et DDU.

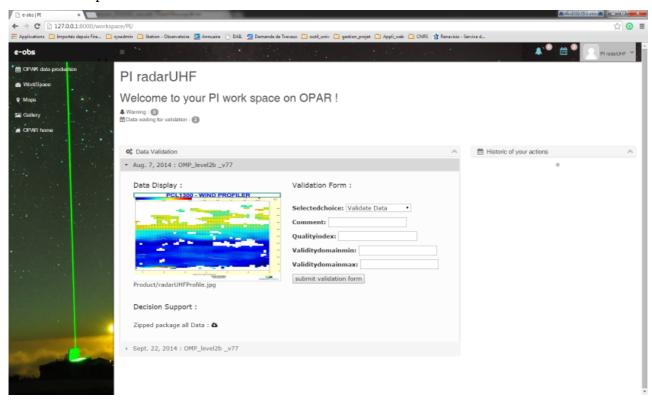
3.5.2 Etat d'avancement

Le développement s'est focalisé sur l'application web jusqu'à aujourd'hui, elle offre une première version des fonctions suivantes :

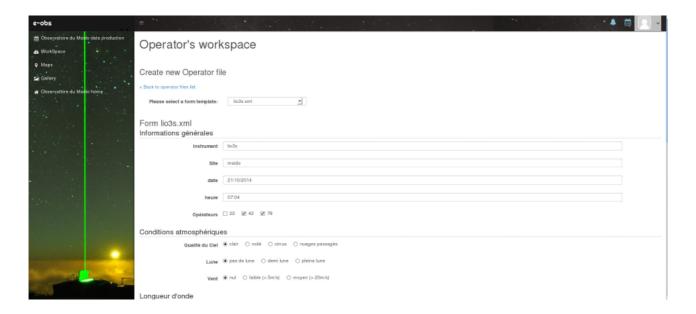
• Visualisation de la production de données stations :



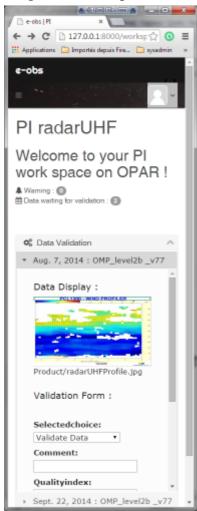
• Un espace de travail PI:

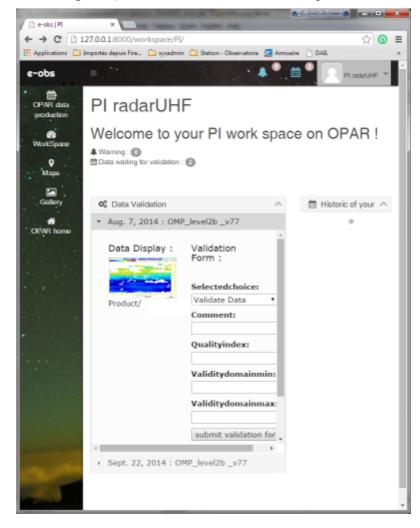


• Un espace de travail Opérateur :



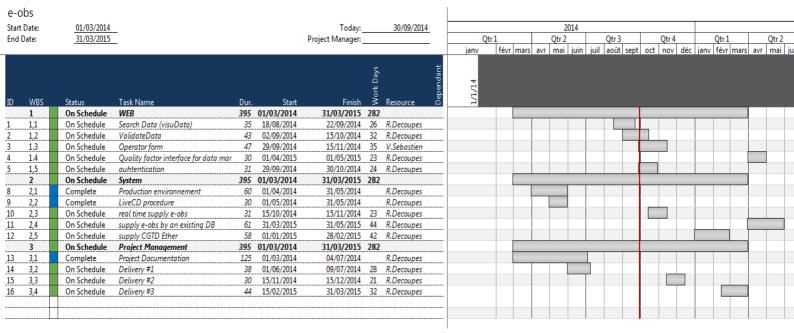
Comme indiqué dans la description du projet e-obs, une place importante a été accordée à l'ergonomie et au design de l'application Web. Le site s'adapte à toute les tailles d'écran (écran large, ordinateur portable, tablette ou smartphone) et se redimensionne automatiquement :





Le niveau technique de la réalisation est satisfaisant. L'utilisation du framework Django et de bootstrap (mise en forme dynamique des pages web) se montre être est un très bon choix dans la pratique. Ces outils offrent une bonne structuration du code qui va favoriser sa maintenance et son appropriation par d'autres équipes. De plus, l'intéraction avec la base données MySql est simplifiée et le rendu visuel est excellent.

Le taux de réalisation de ce développement est cohérent avec le planning initial (voir ci-après le diagramme de Gantt du projet Promise/web au 01/10/2014). La seconde partie du projet va être consacrée aux applications de tâches de fonds (ingestion des données et diffusion automatique des données) et à la consolidation de l'application web.



3.5.3 Priorisation des tâches qui restent à mener

Les prochains développements à venir sont :

- L'alimentation d'e-obs au fil de l'eau.
- La consolidation des espaces de travail opérateur et PI.
- Le développement de l'espace de travail administrateur.
- La diffusion des données élaborées et validées vers le CGTD.

A la fin des deux premiers développements (15/11/2014), une validation complète sera déroulée afin d'effectuer une deuxième livraison d'e-obs. Cette version d'e-obs sera testée auprès d'un PI et des opérateurs instruments à l'OPAR afin de valider son utilisation et commencer à recencer les ajustements souhaitables dans le futur.

En parallèle de la phase de test de cette version sur une filière à l'OPAR, le module d'alimentation du CGTD Ether sera développé (à partir de janvier 2015). Ce module, ainsi que les différents développements critiques qui découlent des remarques, seront developpés de janvier à mi-février afin de mettre à disposition la troisième et dernière version d'e-obs dans le cadre de PROMISE.

3.6 Phase de développement – Promise/arletty : avancement

Cette tâche est menée par l'équipe du CGTD Ether/IPSL. Les outils nécessaires sont prêts à être déployés. La chaîne de production et de livraison doit être mise en service avant mi-novembre 2015 (cf. DR08).

3.7 Statut sur les livrables et la documentation associée

Codes disponibles sur le serveur de développement http://osur.univ-reunion.fr/outils/devspot/.

Tous les documents associés à l'ensemble des logiciels sont en cours d'écriture au fil de l'eau.

Concernant plus particulièrement e-obs :

Iso avec e-obs 0.0.9 déjà livré le 09/07/2014.

Deux autres livraisons d'e-obs sont prévues :

e-obs 0.9.X : 15/12/2014e-obs 1.0.X : 31/03/2015

La version e-obs 0.9.X sera utilisée pour la filière ozone strato à l'OPAR. Le PI et l'équipe technique de l'OPAR vont exploiter le système. Ainsi la version e-obs 1.0.X comportera les développements liés à leurs remarques.

3.8 Rayonnement du projet

Présentation du projet au Workshop du Groupe de Travail Données Rosea&Oraure les 16 et 17 septembre :

- Mise en avant des outils utilisé dans ce projet : GitLab, Python, Django.
- Discussions autour de la définition de l'interface d'échanges station centre de données.
- Discussions autour du format station : NetCDF 4 classic associé à la convention CF.

Discussion en cours avec le SIRTA autour de la structuration de la base de données dans un objectif d'essayer d'être au plus proche afin de favoriser les échanges technologiques et autour de la collaboration sur les développements.

Rédaction d'un document de standardisation autour de la définition des niveaux de traitement, de la standardisation du format et de la nomenclature portée par Franck Gabarrot dans le cadre de ce groupe de travail.

3.9 Interface station-centre de données et standardisation du format des données station

La partie Arletty du projet nous fournit un cas test pour définir l'interface associée au flux centre de données vers station (cf. DR08). Les données 2013 consolidées de la filière lidar ozone stratosphérique 2013 sont prêtes à être livrées. Ce sera là l'occasion de travailler sur la définition de l'interface associée au flux station vers centre de données. La rédaction d'un document compilant une première définition de l'interface centre de données – station a démarré.

Le projet Promise lance l'idée de l'établissement d'un standard pour le format des données produites en station, en particulier pour la partie lidar : Licorne founirait un export standard en NetCDF 4 – CF cohérent avec les attentes du NDACC (templates GEOMS NDACC). Cela rejoint le travail démarré par le SIRTA il y a plusieurs années, qui continue de façon plus large dans le groupe de travail données ROSEA avec la démarche de standardisation s'appuyant sur le NetCDF 4 norme CF. L'équipe de Promise participe à ce travail et, au travers de l'application de cette

standardisation dans Promise, intègre le NDACC-France dans cette démarche.

4. CONLUSION

Le niveau technique de la réalisation est satisfaisant à ce stade, de même que le rayonnement du projet avec les discussions techniques et organisationnelles qu'il entraîne. Le projet est sur la bonne voix mais on note du retard sur la partie lidar ce qui va limiter la réalisation des livrables, en particulier le développement opérationnel des applicatifs d'inversion.

Les premiers mois de réalisation du projet nous montrent que Promise apporte une nouvelle dynamique essentielle dans l'intéraction entre la station et le centre de données. Il ouvre une voix pour la mutualisation des outils au sein de la communauté française, et il pousse à passer un gap organisationnel et technique.